

WHAT IS CLAIMED IS:

1. A computer-implemented method for generating page rankings using a user access log, comprising:
 - 5 extracting implicit links from the user access log;
 - generating an implicit links graph from the extracted implicit links;
 - and
 - computing page rankings using the implicit links graph.
- 10 2. The computer-implemented method of claim 1, further comprising segmenting the user access log into a plurality of different browsing sessions.
3. The computer-implemented method of claim 2, wherein extracting implicit links further comprising using a two-item sequential pattern mining
15 technique to extract the implicit links from the plurality of different browsing session.
4. The computer-implemented method of claim 1, further comprising defining a generative model for the user access log based on the implicit links
20 graph.
5. The computer-implemented method of claim 1, wherein the implicit links graph is a weighted direct graph.
- 25 6. The computer-implemented method of claim 5, wherein the implicit links graph is described by the equation $G' = (V, E')$, where V is a set of vertices representing all pages in a search space and E' encompasses a set of implicit links between the pages.

7. The computer-implemented method of claim 6, wherein each of the implicit links further includes a conditional probability parameter of a page to be visited given a current page.

5 8. The computer-implemented method of claim 1, wherein extracting implicit links further comprises analyzing observed explicit links in the user access log.

10 9. The computer-implemented method of claim 3, wherein the two-item sequential pattern mining technique further comprises moving a gliding window over each explicit link path in the user access log.

15 10. The computer-implemented method of claim 9, further comprising generating order pairs of pages using the gliding window.

11. The computer-implemented method of claim 1, further comprising updating the implicit links graph by setting all weights in two-item sequential patterns to zero.

20 12. The computer-implemented method of claim 11, further comprising adding a support to each of the weights.

25 13. A computer-readable medium having computer-executable instructions for performing the computer-implemented method recited in claim 1.

14. A process for enhancing initial results obtained from a search engine on a computer using a user access log, comprising:
extracting implicit links of pages from the user access log using a two-item sequential pattern mining technique;
30 generating an implicit links graph from the implicit links;
generating two-item sequential patterns from the implicit links;

updating the implicit links graph using the two-item sequential patterns; and
re-ranking the initial search results using the updated implicit links graph.

5

15. The process as set forth in claim 14, further comprising pre-processing the user access log.

10 16. The process as set forth in claim 15, wherein pre-processing further comprises at least one of: (a) data cleaning; (b) session identification; (c) consecutive repetition elimination.

15 17. The process as set forth in claim 15, wherein pre-processing further comprises data cleaning of the user access log by filtering out any access entries for embedded objects.

20 18. The process as set forth in claim 15, wherein pre-processing further comprises performing session identification on the user access log to distinguish each user in the log by their Internet protocol (IP) address.

19. The process as set forth in claim 15, wherein pre-processing further comprises performing consecutive repetition elimination to handle a situation where multiple users have the same IP address.

25 20. The process as set forth in claim 19, further comprising removing IP addresses whose page hits count exceeds a threshold.

30 21. The process as set forth in claim 15, further comprising segmenting the pre-processed user access log into browsing sessions.

22. The process as set forth in claim 21, further comprising grouping consecutive entries of the user access log into a browsing session.

23. The process as set forth in claim 21, further comprising generating
5 ordered pairs of pages from the segmented user access log.

24. The process as set forth in claim 23, further comprising filtering the ordered pairs using a minimum support threshold to remove any ordered pairs that are infrequently occurring in the user access log.
10

25. The process as set forth in claim 14, wherein re-ranking the initial search results using the updated implicit links graph further comprises using a modified implicit link analysis technique for re-ranking.

15 26. The process as set forth in claim 25, wherein the modified implicit link analysis uses a modified re-ranking formula.

27. The process as set forth in claim 26, wherein the modified implicit link analysis uses at least one of: (a) score-based re-ranking technique; (b)
20 order-based re-ranking technique.

28. One or more computer-readable media having computer-readable instructions thereon which, when executed by one or more processors, cause the one or more processors to implement the process of claim 14.
25

29. A computer-readable medium having computer-executable instructions for enhancing local searching of web sites and intranets by mining user access logs, comprising:

30 segmenting the user access log into different browsing sessions;
generating ordered pairs of pages from the browsing sessions to find implicit links;

constructing an implicit links graph from the implicit links;
generating two-item sequential patterns from the ordered pairs;
updating the implicit links graph using the two-item sequential
patterns; and

5 re-ranking search results obtained from a search engine to
enhance the local searching.

30. The computer-readable medium of claim 29, further comprising pre-
processing the user access log using at least one of: (a) data cleaning; (b)
10 browsing session identification; (c) consecutive repetition elimination.

31. The computer-readable medium of claim 29, further comprising
identifying each individual ones of the browsing sessions.

15 32. The computer-readable medium of claim 31, further comprising
identifying in terms of a user identification and a chronological order of pages.

33. The computer-readable medium of claim 29, further comprising
using a gliding window to move over explicit paths of the browsing sessions to
20 generate the ordered pairs.

34. The computer-readable medium of claim 33, further comprising
defining the gliding window size, wherein the size represents a maximum interval
a user clicks between a source page and a target page.

25 35. The computer-readable medium of claim 29, further comprising
filtering the ordered pairs to remove any ordered pairs that are infrequently
occurring.

30 36. The computer-readable medium of claim 35, further comprising
determining a frequency of each of the ordered pairs.

37. The computer-readable medium of claim 36, further comprising:
defining a minimum support threshold; and
applying the minimum support threshold to the frequency of each of
5 the ordered pairs.

38. The computer-readable medium of claim 37, further comprising
discarding an ordered pair if its frequency is below the minimum support
threshold.

10

39. The computer-readable medium of claim 37, further comprising
keeping an ordered pair if its frequency is above the minimum support threshold.

40. A computer-implemented method contained on computer-readable
15 media having computer-executable instructions for execution on a computing
device for enhancing initial search results of a search engine performing a local
search of a web sub-space using a user access log, comprising:

pre-processing the user access log;
segmenting the log into browsing sessions;
20 generating ordered pairs from the browsing sessions;
filtering the ordered pairs using a minimum support threshold to
remove any infrequently occurring ordered pairs to generate two-item sequential
patterns;
updating an implicit links graph using the two-item sequential
25 patterns; and
re-ranking the initial search results using the updated implicit links
graph.

41. The computer-implemented method as set forth in claim 40, further
30 comprising discarding any ordered pairs having a frequency below the minimum
support threshold.

42. The computer-implemented method as set forth in claim 40, further comprising keeping any ordered pairs having a frequency above the minimum support threshold.

5

43. The computer-implemented method as set forth in claim 40, further comprising defining an adjacency matrix to describe the updated implicit links graph.

10

44. The computer-implemented method as set forth in claim 43, further comprising computing a page rank using the adjacency matrix.

15

45. The computer-implemented method as set forth in claim 43, further comprising defining a modified re-ranking formula in terms of the adjacency matrix.

46. The computer-implemented method as set forth in claim 45, further comprising modifying the re-ranking formula using a random walk technique.

20

47. The computer-implemented method as set forth in claim 40, further comprising discarding any ordered pairs having a frequency below the minimum support threshold.

25

48. The computer-implemented method as set forth in claim 47, wherein the random walk technique further comprises a probability parameter.

49. The computer-implemented method as set forth in claim 40, wherein re-ranking further comprises using an order-based re-ranking technique.

50. The computer-implemented method as set forth in claim 49, wherein the order-based re-ranking technique further comprises using a linear combination of page positions contained on two lists.

5 51. The computer-implemented method as set forth in claim 50, wherein one of the two lists is sorted by similarity scores.

52. The computer-implemented method as set forth in claim 50, wherein one of the lists is sorted by PageRank values.

10

53. The computer-implemented method as set forth in claim 40, wherein re-ranking further comprises using an score-based re-ranking technique.

15 54. The computer-implemented method as set forth in claim 53, wherein the score-based re-ranking technique further comprises using a linear combination of a content-based similarity score and a PageRank value of all pages.

20 55. An implicit links search enhancement system for an enhancing initial search results obtained from a search engine by mining a user access log, comprising:

an ordered pairs generator that generates ordered pairs of implicit links from the user access log;

25 an update module that updates an implicit links graph using the ordered pairs; and

a re-ranking module that re-ranks the initial search results based on a modified link analysis technique.

30 56. The implicit links search enhancement system as set forth in claim 55, further comprising a user access log pre-processing module for pre-processing the user access log.

57. The implicit links search enhancement system as set forth in claim 56, wherein the pre-processing module performs at least one of: (a) data cleaning; (b) identification of browsing sessions within the user access log; (c) consecutive repetition elimination.

58. The implicit links search enhancement system as set forth in claim 55, further comprising a user access log segmentation module that segments data in the user access log into individual browsing sessions.

59. The implicit links search enhancement system as set forth in claim 55, further comprising a filter module that removes any infrequently occurring ordered pairs.

60. The implicit links search enhancement system as set forth in claim 55, wherein the a modified link analysis technique includes a modified re-ranking formula and at least one re-ranking technique.

61. The implicit links search enhancement system as set forth in claim 60, wherein the modified re-ranking formula is modified by using a random walk technique and a probability parameter.

62. The implicit links search enhancement system as set forth in claim 60, further comprising an order-based re-ranking technique.

63. The implicit links search enhancement system as set forth in claim 60, further comprising a score-based re-ranking technique.